2020 Census
Program Management Review

Evaluations and Experiments-Related Projects

Kevin Deardorff, Program Manager

8.102 – Administrative Records Fitness-For-Use
8.107 – Administrative Records Modeling
5.101 – Coding, Editing, and Imputation Study
8.105 – Matching Process Improvement
7.101 – Enhancing Demographic Analysis**

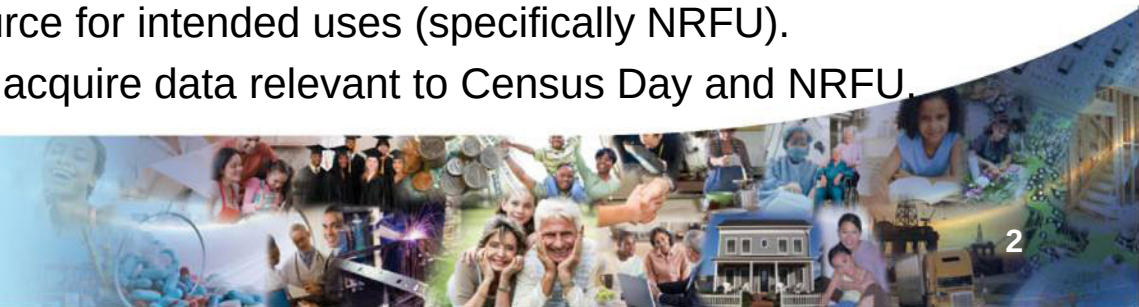
December 14, 2012

WBS 8.102 - Administrative Records Fitness-For-Use

Project Description:

- Investigate administrative records (AR) sources to enhance data collection and processing methods for Nonresponse Followup (NRFU)
  - Explore agreement and disagreement in content across sources to develop approaches to resolve (such as modeling or business rules).
  - Produce recommendations for which AR sources are fit for operational uses such as count imputation, characteristic imputation, and modeling for NRFU.
- Project objectives

1) Describe NRFU addresses with AR available and those lacking AR.
2) Identify AR data that could support count and characteristic imputation.
3) Identify and address conflicting information across AR sources.
4) Assess quality of AR by source for intended uses (specifically NRFU).
5) Assess best time of year to acquire data relevant to Census Day and NRFU.



WBS 8.102 - Administrative Records Fitness-For-Use (Cont.)

Challenges Highlights:

- What is target level of geography and what metrics will gauge AR data quality?
- Can we build AR composites containing different sources for different geographies?
- Can we pursue state-specific (or region-specific or group-specific) data sources to enhance coverage of AR?
- Should we use un-validated person records (without anonymous Protected Identification Keys used for matching)? How would we do so?

WBS 8.102 - Administrative Records Fitness-For-Use (Cont.)

Near-term Project Focus Items:

- Description of NRFU addresses with AR at varying levels of geography.
- Description of AR coverage of addresses responding by proxy.
- Description of AR units that did not match to the Master Address File (MAF) or 2010 Census.
- Evaluation of using AR at block level instead of housing unit level.
- Description of which sources are best when combined and which sources could be excluded from models depending on intended use.
- Use of Census Coverage Measurement (CCM) data to evaluate person records with conflicting address data.
- Evaluation and improvement of "best address" selection when the same person is observed in multiple places.

WBS 8.107 - Administrative Records Modeling

Project Description:

- Consolidate and conduct research efforts for examining optimal scenarios of curtailing contacts during the Nonresponse Followup (NRFU) field operations with the supplemental use of administrative record sources.
- Project Objectives

1) Simulate how administrative sources can be used to supplement curtailed NRFU data collection.
   - Administrative records usage based on either a rule-based algorithm, model or both
   - Count and characteristic imputation approaches that can possibly include the use of administrative records sources

2) Assess the coverage and cost of simulated scenarios.

3) Compare resulting scenarios to identify possible combinations of NRFU visits and administrative records use to pursue.

WBS 8.107 - Administrative Records Modeling (Cont.)

Project Description (continued):

- Coordinate with the 8.102 Administrative Records Fitness-for-Use team based on their recommendations.
- Coordinate with the 5.101 Coding, Editing and Imputation team.
  - 8.107: research involving curtailed NRFU, imputation, and administrative records is focused on the next 6 months
  - 5.101: long-term focus on methods to improve count imputation and characteristic imputation, with and without the use of administrative records
- Coordinate with other teams researching field operation and budget implications.

WBS 8.107 - Administrative Records Modeling (Cont.)

Challenge Highlights:

- Can we have NRFU contact strategies where certain areas or housing units would receive very few or no field visits?
- Can we have NRFU contact strategies where different areas or housing units receive different number of visits?
- What are the coverage metrics necessary to assess simulations?

WBS 8.107 - Administrative Records Modeling (Cont.)

Recent Accomplishments/Near-Term Project Focus:

- Shared initial research of curtailed NRFU contact strategies using both count imputation methods without administrative records and approaches that used administrative records.
- Continue researching ways of using administrative records sources to supplement curtailed NRFU strategies.
- Utilizing count and characteristic imputation methods to account for remaining census missing data.
- Determining metrics to assess the accuracy of a scenario of NRFU contact strategy and administrative record source usage.
- Estimate approximate cost savings for different NRFU contact strategies proposed.

WBS 5.101 - Coding, Editing, and Imputation

Project Description:

Develop improved methods for missing data in the census:

- Count imputation
  - Unit status of address: occupied, vacant, or nonexistent?
  - Number of persons in housing unit or group quarters.
- Characteristic imputation, editing, and coding
  - Population characteristics: race, Hispanic origin; age, sex, relationship; tenure (own or rent)
  - Housing characteristics: status (type) of vacant units, building type (single unit, multi-unit, trailer/boat, other)

WBS 5.101 - Coding, Editing, and Imputation (Cont.)

Project Objectives:

- Improve methods for
  - count imputation: assign status of address, and, if necessary, number of people in house
  - characteristic imputation
  - editing characteristic data
  - coding for write-ins of race and Hispanic origin

- Consider approaches that
  - include the use of available administrative records
  - enhance current procedures *without* using admin. rec.

WBS 5.101 - Coding, Editing, and Imputation (Cont.)

Recent Accomplishments (1):

- Under some "curtailed NRFU" scenarios (e.g., limit nonresponse follow-up to fewer visits), examined
  - resulting 2010 count imputation rates
  - effect on the total census count
- Studied results from modeling count imputation
  - using prior Census Bureau data
  - using public and private sources of administrative records

5.101 - Coding, Editing, and Imputation (Cont.)

Recent Accomplishments (2):

- For imputation and editing, investigating alternative statistical and software approaches.
  - studying more flexible, adaptable statistical methods
  - designed new system (by CSRM) to handle new approach
- For coding, *continuous flow* statistical quality control.
  - designed preliminary methodology
  - simulated some positive results, relative to batch approach
  - refined initial draft of specs for continuous flow system

5.101 - Coding, Editing, and Imputation (Cont.)

Risk Highlights:

- If research assumptions differ from conditions in the 2020 Census, the results may not be valid, e.g.,
  - Content options, e.g., combined race/Hispanic
  - Who responds under new collection mode(s)
  - Policy related to administrative records
  - Reference period of admin. records vs. census day

5.101 - Coding, Editing, and Imputation (Cont.)

Challenges:

- Timeliness of policy decisions.
- The ability to integrate software in real time (at the door; while responding via Internet) or near real time (over night).

5.101 - Coding, Editing, and Imputation (Cont.)

Near-Term Project Focus Items:

- Continue developing new software.
- Study and understand the structure of administrative records.
- Incorporate administrative records into the research.
- Develop appropriate statistical metrics to evaluate results from alternatives; consult with other teams.
- Consider changes in requirements due to possible curtailed NRFU scenarios.

WBS 8.105 – Matching Process Improvement

Project Description:

- Research and evaluate the methodology, techniques, and technology to improve matching of addresses and persons for the 2020 Census.
  - Research and assess matching techniques (including rule-based and probabilistic) to identify optimal methods by application.
  - Determine how to select cutoffs for probabilistic matching to optimize each application (i.e., how much agreement is needed between records to determine that they are actually a match).
  - Research error measurement in matching techniques, including determination of error and assumptions about causes.
  - Research and evaluate software to perform standardization/ preprocessing of data.
  - Identify what data are needed to improve the quality of matching.

WBS 8.105 - Matching Process Improvement (Cont.)

Recent Accomplishments/Near-term Project Focus (1):

- Compare address standardization options using evaluation criteria and test decks identified by the team.
- Continue researching refinements to current methods used to identify person duplication in the census (e.g., modifying measures of agreement based on different measures of geographic distance).
- Identify refinements needed to current record linkage and error measurement methods.
- Evaluate matching cutoffs by reviewing 2010 Census Coverage Measurement clerical results and research using methods to automate the determination of matching cutoffs.
- Refine objectives to include comparisons of the current system used to identify census duplication and the system used to link administrative records for the 2010 Census Match Study.

WBS 8.105 – Matching Process Improvement (Cont.)

Recent Accomplishments/Near-term Project Focus (2):

- Leveraging results from the Geographic Support System Initiative (GSS-I) Address Standardizer Evaluation.
- Leveraging GSS-I efforts for interactive/real-time address matching and geocoding.

Challenge Highlights:

- To establish a robust infrastructure for testing new and enhanced methods, we are defining testing methodologies in order to best predict infrastructure needs.